DOCUMENT RESUME

ED 360 933 HE 026 671

AUTHOR Keating, Jean C.
TITLE SCHEV Integrated Data Base: The First Step. The SCHEV Student Data Base. AIR 1993 Annual Forum Paper.
PUB DATE May 93
NOTE 14p.; Paper presented at the Annual Forum of the Association for Institutional Research (33rd, Chicago, IL, May 16-19, 1993).
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Admission; *Database Design; *Databases; Data Collection; Data Processing; Degrees (Academic); Enrollment; Higher Education; Information Needs; Private Colleges; Public Colleges; Student Financial Aid
IDENTIFIERS *AIR Forum; *State Council of Higher Education for Virginia

ABSTRACT

This paper reports on the implementation of the first two parts of an integrated data base designed to replace the fragmented surveys and data files of Virginia's institutions of higher education. The two parts being examined involve the State Council of Higher Education for Virginia (SCHEV) Student Data Base and the Institutional Information File. The SCHEV Institutional Information File will provide institutional specific information used in checking the various data bases. The design of this file allows expansion of information through the addition of different types or groups of records. The 1992 file consists of three groups of records: institutional specific information, one per institution; multiple records per institution, one record per student level per institution; and multiple records per institution, one record per degree per degree level per institution. A file reflecting all courses taught during a year at each institution is in the process of being added to the database. The SCHEV Student Data Base will consist of five data files: (1) fall headcount, (2) annual course enrollment, (3) financial aid, (4) degrees conferred, and (5) admission. First, the report describes each part's data file design, followed by a section describing the implementation phases for 1992 through 1994. It concludes with a progress update. (GLR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 360 933

SCHEV Integrated Data Base: The First Step
The SCHEV Student Data Base

Jean C. Keating

HE026 621

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.
☑ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

AIR

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

for Management Research, Policy Analysis, and Planning

This paper was presented at the Thirty-Third Annual Forum of the Association for Institutional Research held at the Chicago Marriott Downtown, Chicago, Illinois, May 16-19, 1993. This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of Forum Papers.

Jean Endo
Chair and Editor
Forum Publications
Editorial Advisory Committee

SCHEV Integrated Data Base: The First Step
The SCHEV Student Data Base

Jean C. Keating

INTRODUCTION

Through the late '70s, data collection efforts of the State Council of Higher Education for Virginia (SCHEV) relied on surveys to provide needed information. The '80s saw the addition of more and more individual record collections in specific areas: permanent employee information at public institutions were covered by the PMIS system, facilities at institutions expanded from room-by-room facilities inventory to utilization reports, student specific reporting for assessment and financial aid as well as special reports for retention and affirmative action were added to the surveys already in place.

In 1991, SCHEV acquired a RISC/6000 computer and began the design and implementation of an integrated data base to replace the fragmented surveys and data files. This integrated data base will eventually cover students, degrees, financial aid, assessment, faculty, facilities and finance. Implementation of the first two parts of this integrated data base are underway. These two parts are the SCHEV Student Data Base and the Institutional Information file.

INSTITUTIONAL INFORMATION FILE

An Institutional Information data file provides institutional specific information used in checking the various data bases. Files specifications for the first three record types of this data file were finalized in June 1991 and submitted now for the first time in the fall of 1992.

This file was not subject to the extremely tight security of the student headcount file also implemented in fall 1992. Virginia's institutional staff who were using electronic transmission procedures for the first time felt comfortable with this file as the trial effort in electronic transmissions.

The design of this data file allows expansion of information through the addition of different types or groups of records. For the 1992 submission, this data file consisted of three groups of records:

Type 1 record: Institutional specific information, one per institution.

Type 2 records: Multiple records per institution, one record per student level per institution.

Type 3 records: Multiple records per institution, one record per degree per degree level per institution.

Over the summer type 4 records for this file will be added to reflect all courses taught during a year at each institution. There II-type 4 records will link the Annual Course Enrollment data file of the Student Data Base with faculty efforts and facilities data bases to come.

Student Data Base

The SCHEV Student Data Base was developed by the Research Section of SCHEV in cooperation with Virginia's institutions of higher education. Virginia's eighty-four state-supported, independent not-for-profit and independent for-profit institutions participate in this reporting effort.

Routine reporting efforts by Virginia's institutions have been characterized for more than a decade by a close working relationship between SCHEV staff and one individual at each institution know as the **reports coordinator.** This individual at each institutions is named yearly by the institution's president and acts as the contact point on all reporting efforts to SCHEV. Not surprisingly, at institutions with Institutional Research offices, this coordinator is usually the Director of Institutional Research.

Twenty-one of these reports coordinators form a subcommittee call the Reports Advisory Committee (RAC), which has worked extensively with SCHEV staff over the past year to develop and refine a student data base that considers **when** and **where** data are available within the institution as well as what data are needed. The results of these mutual efforts are contained in the **SCHEV DATA DICTIONARY** and the **SCHEV Student Data Base: Record Layouts.** [If you would like your own copy of either or both of these documents, a sheet is provided at the end of this paper for your use in requesting them. The documents weight entirely too much for multiple copies to have accompanied me to Chicago.]

The SCHEV Student Data Base consists of five data files which are:

* **Fall Headcount Data File:** - this file follows where possible the format of the Uniform Student Data System

(USDA) master file in use by many public institution since 1980 although some data elements are added; it will be used to generate the federal and state fall headcount reports; it will be collected from institutions beginning in November, 1992 and will replace six headcount survey reports (SCHEV B2, SCHEV B3, SCHEV EF, SCHEV R1, SCHEV RS, SCHEV B5). [Individuals who desire more information on the USDA system used in Virginia prior to the SCHEV Student Data Base may use the enclosed request form to obtain a copy of the description of USDA given in the 1984 CAUSE proceedings.]

* **Annual Course Enrollment Data File:** - this file will be a new reporting endeavor for the private institutions, but is consistent with data prepared, since 1975, by state-supported institutions as part of the Student Data Module (SAM). This file will allow the generation of full-time equivalent enrollments for all institutions not just for state-supported ones. Several private institutions have asked us to develop a measure other than fall headcount to define size and breadth of institutional efforts. The Annual Course Enrollment file expands the information available from state-supported institutions to show courses taken by senior citizens and grades received by transfer undergraduate students by course, the latter providing one of the assessment reporting requirements. This file will be implemented in July, 1993 and will replace the SAM.

* **Financial Aid Data File:** - this file is an expansion of two existing files. Student-specific financial aid data have been collected since 1985 on the SCHEV S6 (state-supported) and since 1987 on the SCHEV S7 (independent) report files. Several years ago, we began a process to consolidate these reports into one (tentatively labeled the SCHEV S report). This consolidated report was scheduled to begin with the 1991-92 academic year. The decentralization of the TAG and the implementation of these data as part of the Student Data Base have pushed back the start-up of this report to the 1992-93 academic year. Some data elements are new, but most were on either the SCHEV S6 or the SCHEV S7. Reporting requirements for this data file, implemented in September, 1993 to reflect the 1992-93 financial aid received, require social security number on all students included in this and in all other files of the data base over all years the student is enrolled. Location of domicile is one of the new data elements required on the Financial Aid Data File, enabling SCHEV to determine distribution of recipients by planning district, city/county, etc. This data file replaces the SCHEV S6 and SCHEV S7 reports.

* **Degrees Conferred Data File:** - this file contains a subset of the elements from the USDA. It replaces the IPEDS C1/C2 reports and collects considerably more information regarding degrees and awards given by Virginia's institutions, including race, gender, and location of domicile of recipients by the exact degree attained (by six-digit CIP). The data file will be collected in September of 1993 to cover the 1992-93

academic year.

- **Admission Data File:** - this file will be the most challenging for the institutions to automate. Data elements for this file are not currently covered under USDA. An admissions data file will be submitted twice during the year. A Fall Admissions Data File will be implemented November 1993 to reflect summer and fall admissions for 1993. An annual Admissions Data File will be collected in June, 1994 to reflect admissions not shown on the fall file. The Admissions Data File for state-supported institutions consists of three parts. Part 1 consists of the basic information now submitted in aggregate form on the SCHEV B8. Part 2 contains supplemental information necessary to replace the SCHEV HSR1 report which is the assessment component that measures performances of first-year freshmen from Virginia high schools. Part 3 contains supplemental information on transfer-undergraduate performance and is required of four-year public institutions and Richard Bland College. The Fall Admissions Data File data files replaces the SCHEV B8; Fall and Annual Admissions Data File replace the SCHEV HSR1 reports.

SCHEDULE FOR 1992-94

A detailed description of the stages in the development of data files are summarized by number below. The phases of development and deadlines for completions of segments of the Institutional Information file and the Student Data Base are given by a graphic display attached. A brief description of the efforts involved in each phase are shown below.

Phase 1 From initial planning to completion of this phase, this stage offers the maximum potential for input by SCHEV staff. On the graphic, this phase (as well as phases 2 and 3) are denoted by an open square. Activities from initiation of work to completion of this phase are centered at SCHEV.

The Research staff review existing surveys and/or data files relating to subject area, noting needed data elements and time frame, comparing extent of higher education already covered with current and expected future needs. Existing data collections are compared with a four-year log of requests for information in subject area to develop draft of data needed; data files needed are defined, time frame covered by each and submission times determined; data elements are defined,

record layout, and data dictionary pages for each data element in the record are developed.

Data dictionary contains codes, definition of data element, standard name, length and format, requirements for private institutions or subset of public if need for this element differs across the types of institutions, specifies purpose for which collected, etc. At this point, the existence of future edit criteria in each of four levels is noted though not precisely defined. The edit levels are: A, those pertaining to the element alone; B, those pertaining to the element in relation to other elements on the same record; C, those comparing the element to information supplied in the Institutional Information file; and/or D, those comparing the data element to elements on other records.

Phase 2 This second phase of the development of the data base/data files involves the translation of decisions concluded with Phase 1 completion into computer instructions. Like Phase 1 it is noted on the graphic version of the schedule with an open square. As this phase progresses, changes create more work than would have been the case if these same changes were dealt with in phase 1. Phase 2 ends with a draft version of data element lists, record layouts and data dictionary pages developed and distributed to Reports Advisory Committee (RAC), and reviewed with RAC for comments, additions and deletions

Phase 3 This third phase of the development ends with a revise record layout and data dictionary pages consistent with RAC discussions being distributed to all institutions for use. Like phases 1 and 2, SCHEV staff may still make changes at this stage, though such changes now are much more difficult to deal with since a recycling back through RAC or through Phase 2 would be necessary.

Phase 4 Develop update, edit and display programs which are used to screen the loading of institutional data submissions, the editing of these data submissions and the display programs which denote errors or acceptance of information; load final version to open access area for use by institutions and/or by

Research staff; request test data from institutions and utilize actual data in testing this phase

The end of this fourth phase is denoted by a darkened square on the graph; changes at this stage should be limited to minor extensions of codes but not data elements. Any changes at this stage of development are costly.

Phase 5 Add error codes to data dictionary for each data element and distribute final version of data dictionary pages to all institutions; institutions submit data based on instructions from this phase of the development

Phase 6 Develop SAS programs to extract aggregated data from unit-record data files, append aggregate data for most recent term/year to SAS Historical Data Files (for B2/B3, EF, R1, RS, B5, SAM and C1/C2). This stage is indicated on the schedule with a open circle symbol. It means that input from staff and institutions are -- for the most part -- locked out. Institutions will submit data files based on Phase 5 instructions. While Research staff can and will develop more extraction programs against the data file described in phase 5, it's too late to decide a different split on codes or data elements are needed except for adding them the next cycle.

HOW WE'RE DOING SO FAR

Virginia's institutions were encouraged to submit the Institutional Information data file and the Fall 1992 Headcount Data File electronically. Upon request, each institution was given an account on the SCHEV RS6000 for their use and provided with instructions on how to get on and transmit a file. Edit programs for these two files were provided for their use in editing the files before releasing them to SCHEV's research group.

Fifteen state-supported institutions, 16 independent not-for-profit, and four independent for-profit institutions loaded their files to the SCHEV computer electronically, edited them on our system, corrected their data files and released clean data files to SCHEV. The remaining institutions, with two exceptions, submitted

ASCII files on tape or diskette. Two institutions, with a combined enrollment of 110 students could not produce machine-readable data and submitted hard-copy reports which had to be keyed at SCHEV. Of the total 350,602 records in the fall 1992 headcount enrollment file, only 110 had to be keyed at SCHEV.

In addition to edit programs provided for institutions' use, a series of display programs which replicated the familiar headcount reports (SCHEV B2: Off-campus enrollments, SCHEV B3: On-campus enrollments, SCHEV R1: Enrollments by City/County, SCHEV RS (and IPEDS EF Part C: Enrollments by state of domicile, SCHEV B5 and IPEDS EF Part B: Enrollments by age and the IPEDS EF Part A) could be run by each institution.

The edit programs used with the two files submitted in fall 1992 two levels of responses: 1) an error summary which identifies groups of problems but does not print the record identification which is student social security number on the fall headcount file and 2) a detail error listing which does. When SCHEV staff performed the edits, only the first of these two levels of errors were returned to the institutions. When resolutions of problems required the second, detailed level, resolutions of problems were done by phone.

When an institution's data submission was accepted by SCHEV as error free, SCHEV staff reran displays of the familiar headcount survey forms and returned them to the reports coordinator as confirmation of SCHEV's acceptance of the submission and for final validation of the data.

The edits on these first files of the SCHEV Student Data Base are

1. **tight;** edit criteria conform strictly to definitions described in the SCHEV Reporting Guidelines, a document produces almost eight years ago with the help of the Reports Advisory Committee and updated yearly. One institution with an enrollment of approximately 8,000 got 27,840 errors on their first submission.

2. **not fully implemented;** the Data Dictionary lists four different levels of edit criteria, the fourth of which edits a data element in relationship to elements in other files of the data base which haven't yet been submitted.

3. **not tight enough:** we are now writing more extensive checks for longitudinal changes from year to year on the total file to avoid problems encountered with these first submissions on incomplete file submissions and illogical (though technically correct) submissions.

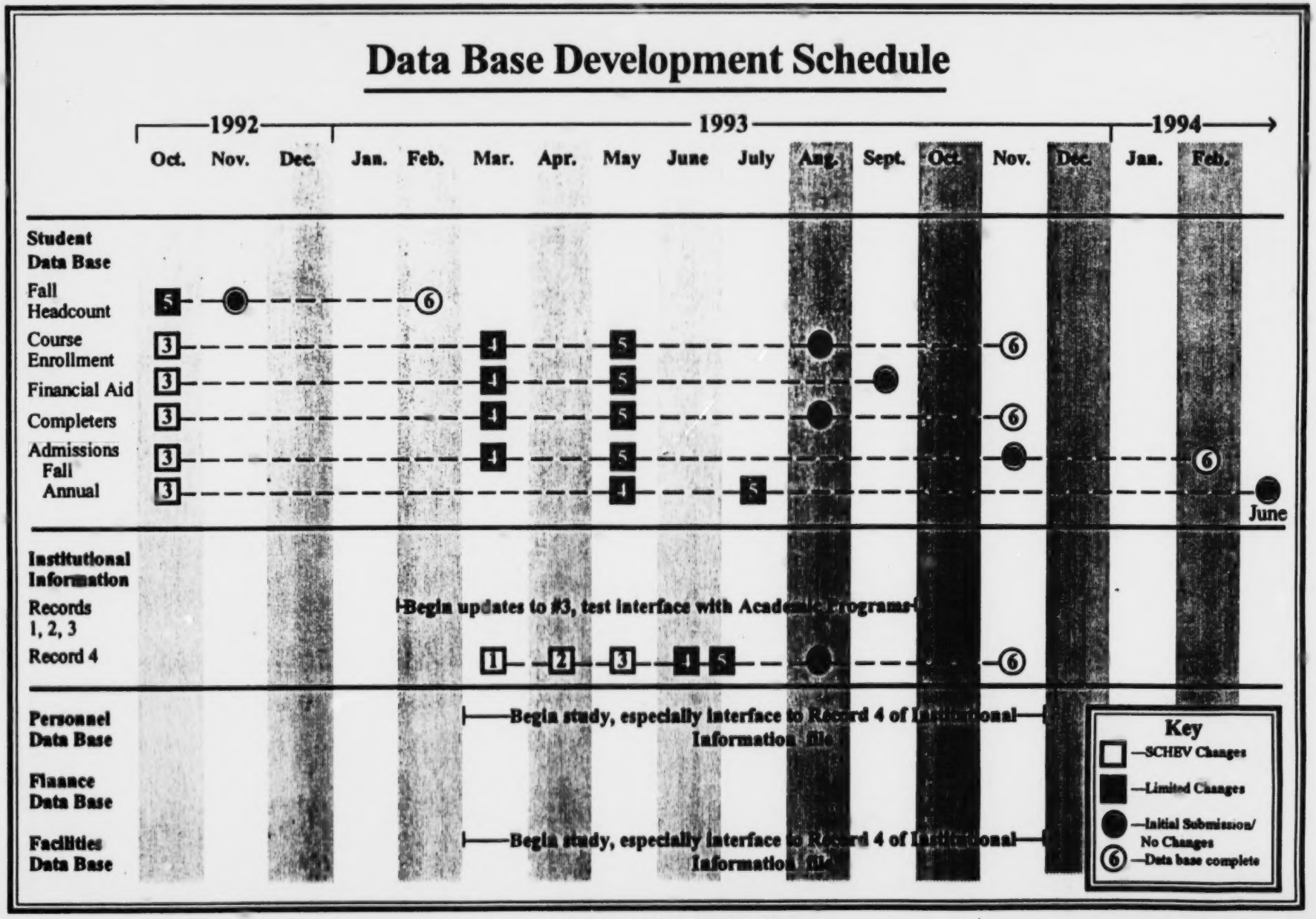
The new system obviously imposes a substantially increased

reporting burden on Virginia's institutions. But the product offers a wonderful opportunity for access to useful and accurate data for the institutions as well as SCHEV. Since 1991, SCHEV has published a listing of aggregate data files available for institutional use which provide headcount and completers data on all Virginia institutions back to 1978. As a first step in supporting the institutional research efforts at Virginia's institutions, the fall 1992 headcount data files were aggregated into the surveys they replaced. Access to these aggregate files is open to any Virginia institution using the SCHEV computer.

As more institutions utilize these aggregate files based on the SCHEV Student Data Base, SCHEV will add other aggregate files of information newly available via the Student Data Base - the number of 18 to 21 year old college students by planning district, for example. SAS programs that display data from all aggregate file will be expanded; programs written by SCHEV or contributed by institutional staff will be shared for all of Virginia's institutions to use in extracting information.

When technological advancements permit the use of one-way screens to guard student security while allowing free access to the raw data files, the SCHEV Student Data Base will serve as central pool of information for research efforts by all of Virginia's institutions.

Data Base Development Schedule



BEST COPY AVAILABLE

DOCUMENTS AVAILABLE FROM SCHEV

Documentation referenced in this paper are available from the State Council of Higher Education for Virginia upon request. Please check below to indicate your interest in receiving a copy of the referenced material and return this completed form to:

Jean C. Keating
SCHEV - James Monroe Building, 10th. floor
101 North 14th. Street
Richmond, Virginia 23219

_____ **Data Dictionary: SCHEV Student Data Base**

_____ **Record Layouts: SCHEV Student Data Base**

_____ **SCHEV Reporting Guidelines: 1992-93**

_____ **SCHEV Historical Data Sets; April 1993**

_____ Carter, Fletcher and Keating, Jean C.: **A Statewide Reporting System Pays Off: The Virginia Experience;** Proceedings of the 1984 CAUSE National Conference, December 4-7, 1984.

NAME AND ADDRESS OF REQUESTOR:

Name: _____

Title: _____

Telephone number: _____

Institution name and mailing address:

_____

_____

_____